Watermarking of MPEG-4 Videos

Abhinav Gupta and Phalguni Gupta

Department of Computer Science and Engineering
Indian Institute of Technology, Kanpur, India - 208016
{abhigupt,pg}@cse.iitk.ac.in

Abstract. A MPEG-4 compressed domain video watermarking method is proposed and its performance is studied at video bit rates ranging from 64 Kb/s to 900 Kb/s. The watermark is inserted by modifying Discrete Cosine Transformation (DCT) coefficients. The strength of watermark is changed according to local frame characteristics to reduce impact on visual quality. The algorithm's performance is also studied for watermarking bits in a frame ranging from 1 Kb/frame to 3 Kb/frame. The watermark is attack-free against attacks like scaling, rotation and cropping even if blind-techniques are used.

1 Introduction

Digital Watermarking is imperceptible insertion of information into multimedia data which act as a signature in the video. The classical approach to watermarking of a compressed video stream is to decompress the video and use a spatial-domain or transform-domain watermarking technique and then recompress the watermarked video. Some of the disadvantages of this approach are: noise addition due to recompression, bad quality due to no knowledge of compression parameters results and computational complexity. Another approach is to insert the watermark in compressed domain only. In this approach the watermark is inserted in some syntactical elements like DCT coefficients of a partially decoded video.

Hartung [1, 2] describes techniques to embed a spread-spectrum watermark into MPEG-2 [3] compressed videos as well as into uncompressed video. For compressed domain watermarking they decode the video to obtain the DCT coefficients of each frame and insert the watermark by modifying those DCT coefficients. Langelaar[4] describes a compressed domain watermarking technique called Differential Energy Watermark (DEW) in which the video is partitioned into groups of blocks each of which is further divided into two sets of equal size as determined by the watermark embedding key. By comparing the energy of selected DCT coefficients within two sets, a single payload bit is expressed. Jordan et al. [5] proposed a method for the watermarking of compressed video that embeds information in the motion vectors. Hsu and Wu present a method for watermarking [6] which modifies middle frequency DCT coefficients in relation to spatially and temporally neighbouring blocks. Nicholson [7] evaluated the watermark robustness and video quality after video is watermarked and compressed

D. Zhang and A.K. Jain (Eds.): ICBA 2004, LNCS 3072, pp. 746–752, 2004.

with MPEG-4. However none of these techniques addressed direct watermarking of these MPEG-4 videos.

In this paper, a new compressed domain watermarking technique for MPEG-4 [8] video streams is presented. The approach is similar to [1] as watermark is inserted in DCT coefficients. However use of synchronization templates makes it robust and local gain method improves the quality.

In Sect. 2 an overview of MPEG-4 technique is presented. This is followed by the proposed method and results in Sect. 3 and Sect. 4 respectively. Section 5 proposes a protocol for subjective evaluation of the system and the paper concludes in Sect. 6.

2 Overview of MPEG-4

MPEG-4 [8] encodes the visual information as objects (natural, synthetic video and still textures). MPEG-4 encodes a description of the scene for proper rendering of all objects. An MPEG-4 visual scene may consist of one or more video objects. Each video object is characterized by temporal and spatial information in the form of shape, motion and texture and corresponds to a 2D object in the scene. A Video Object Plane(VOP) is a time sample of video object. VOP's can be encoded independent of each other or dependent on each other by motion compensation. A VOP contains the encoded video data in form of macroblocks. A macroblock contains a section of the luminance component and the spatially subsampled chrominance components. In MPEG-4 visual standard there is support for only one chrominance format for a macroblock, the 4:2:0 format. In this format, each macroblock contains 4 luminance blocks, and 2 chrominance blocks. Each block contains 8x8 pixels encoded using DCT transformation. The DCT coefficients are then adaptively quantized to achieve low bit rates.

3 Proposed Method

In the proposed method, a watermark signal is inserted directly into MPEG-4 compressed bit-stream while detection is performed using compressed bit-stream without any watermark signal. Section 3.1 discusses the formation of watermark signal from the original message signal and synchronization templates. Section 3.2 addresses the process by which the watermark signal is embedded in the MPEG-4 videos. Section 3.3 discusses the local adaptive gain method to increase quality of the videos.

3.1 Spread Spectrum Watermark Signal

The watermark signal is often limited to a small value to ensure the imperceptibility and subject to interference from the host signal and additional noise arising from subsequent processing. A spread spectrum signal is vulnerable to synchronization errors occurring after scaling, cropping and rotation. A pair of templates is imposed on the spread spectrum signal to combat synchronization losses. The first template restricts the watermark signal to have a regular periodic structure. In particular, the watermark signal $w(x, y)$ is constructed by repeating an elementary watermark tile $\hat{w}(x, y)$ in a non-overlapping fashion. If the tiling is done properly the peak always occur at the centre of each tile. If a linear transformation A is applied to watermarked VOP, the autocorrelation coefficients $h(x, y)$, thus the new peaks move to x' and y' according to

$$[x' \ y']^T = A[x \ y]^T \tag{1}$$

The second synchronization template forces $w(x, y)$ to contain a constellation of peaks in frequency domain. This requirement is met by making $\hat{w}(x, y)$ as combination of the message bearing symbol $m(x, y)$ and the synchronization signal $g(x, y)$. In frequency domain this $g(x, y)$ contains peaks in mid frequency band each peak occupying one frequency coefficient and having unity magnitude. After the geometrical transformation A applied to image the FFT coefficient $F(u, v)$ move to a new location (u', v') according to the equation

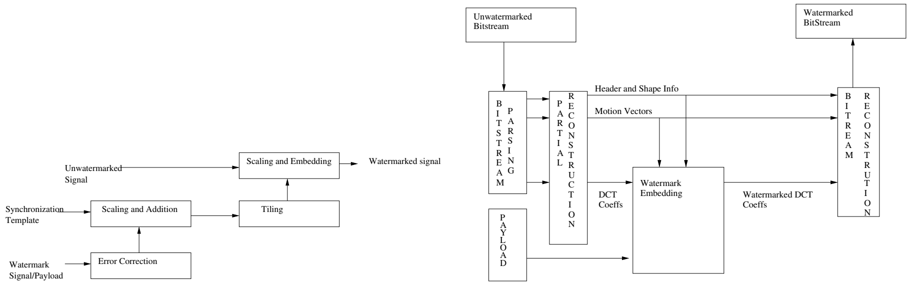
$$[u' \ v']^T = (A^T)^{-1}[u \ v]^T \tag{2}$$

Thus using the two equations above we can obtain the linear transformation A applied to the image and hence revert the transformation to perform the detection method. Figure 1(a) outlines the process of creation of watermark signal.

3.2 Watermark Embedding

This section describes embedding the watermark directly to the bit-stream generated in accordance with the Advanced Simple Profile (ASP) of the MPEG-4 standard. The watermark signal $w(x, y)$ is added to the luminance plane of the VOPs. Since the DCT is a linear transform adding the transformed watermark signal directly to DCT coefficients of the luminance blocks is equivalent to addition in spatial domain. An elementary bit-stream is parsed down to the block level and variable length coded motion vector and DCT coefficients are obtained. Motion vectors are reconstructed using VLC decoding and reversing the prediction steps wherever applicable. After the watermark signal is embedded, VLC codes are regenerated and bit-stream is reconstructed. Fig. 1(b) depicts the whole outline.

Since a 96x96 signal is embedded in a 192x192 image, a 8x8 block is embedded in 16x16 block. Hence, the total number of permutations possible are P_{64}^{256}. This property helps to introduce the Digital License Number. Each authorized user has a key that maps to one of the configuration of all possible permutations. Quadratic Chaining is a method where iteration over $(i + i^2)K mod 256$, where K is the license number, yields 64 places to insert watermark. Hash Table is another approach to do the same.



(a) (b)

Fig. 1. a) Schematic Diagram for preparation of watermark signal. b)Outline for a watermark embedder. The embedder mimics the MPEG-4 decoder

3.3 Adaptive Local Gain Methodology

The adaptive local gain methodology improves the performance of the watermark. For relatively smooth regions of the video, where even a small amount of distortion may be visible , the local gain control reduces the watermark embedding power to minimize the watermark perceptibility. For relatively busy or textured regions of the image, the local gain control increases the embedding power for improved robustness. The local gain method uses a local activity measure to adjust the watermark power on a block by block basis, which is obtained directly from DCT coefficients for intra-blocks and predicted using motion vector information for predicted blocks. The gain model outputs a local gain $L(x, y)$. The watermark coefficients are then weighted by $L(x, y)$ to produce the watermark signal that will be embedded into video:

$$W^*(x, y) = \alpha L(x, y) W(x, y) \tag{3}$$

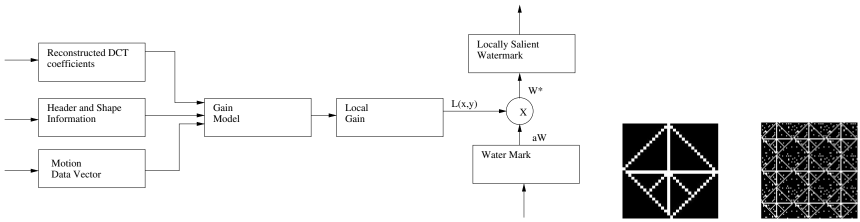
where W* is the watermark that will be embedded, α is user selected global gain and W is the watermark signal prior to gain adjustment. For each VOP, local gain weights are decided based on the estimated activity in the VOP. For Intracoded-VOP's the $L(x, y)$ is

$$L(x, y) = DCT(f(x), g(y))^2 \Sigma DCT(i, j)^2 \tag{4}$$

where $f(x)$ and $g(y)$ map the $(x, y)th$ pixel to $(f(x), g(y))$ DCT coefficient. For predicted VOPs we used the same formula but instead of Total Energy being $\Sigma DCT(i, j)^2$ it is taken to be

$$A_i = \frac{N1}{N} A_A + \frac{N2}{N} A_B + \frac{N3}{N} A_C + \frac{N4}{N} A_D \tag{5}$$

where N1,N2,N3 and N4 are the number of pixels that moved from Area A,B,C,D to the block currently being considered. The sketch for local adaptive gain is shown in Fig. 2(a).



(a) (b) (c)

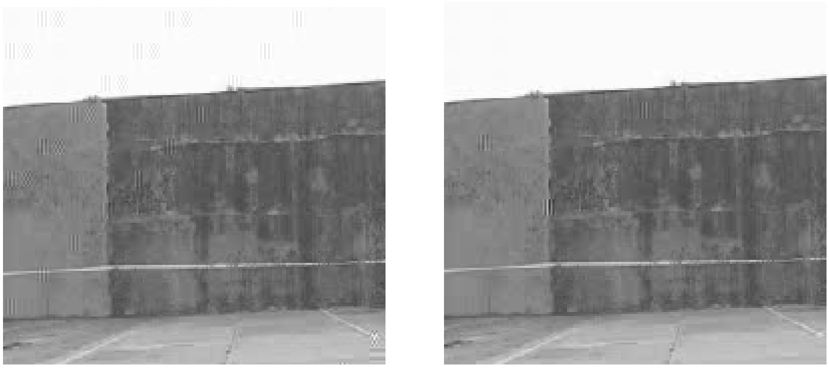
Fig. 2. a) The sketch for Local Adaptive Gain Method. b) The Original watermark signal that is to be inserted. c) The final watermark signal after tiling

(a) Video I(900 Kb/sec) (b) Video II(64 Kb/sec)

Fig. 3. a) Watermarked Video of bitrates 900 Kb/sec and payload bit-rate of 1 Kb/frame. b) Watermarked Video of bitrates 64 Kb/sec and payload bit-rate of 1 Kb/frame

4 Results

The algorithm has been tried with the two videos (one having lot of texture compared to other) taken in IIT Kanpur as there is no standard video dataset for watermarking approaches. The videos were compressed at different bit rates (64-900 Kb/sec) and the watermark signal bit rates were varied from 1 kb/frame to 3 Kb/frame. The frames were 192 x 192 in dimensions and a watermark of 32 x 32 was inserted. The preparation of watermark signal has been shown in figures. 2(b), 2(c). The results of watermark embedding at 1 Kb/frame and 900 kb/s and 64 kb/s videos are shown in Fig. 3. Figure 4 shows the performance of local adaptive gain module.



(a) No Local Adaptive Gain (b) Local Adaptive Gain

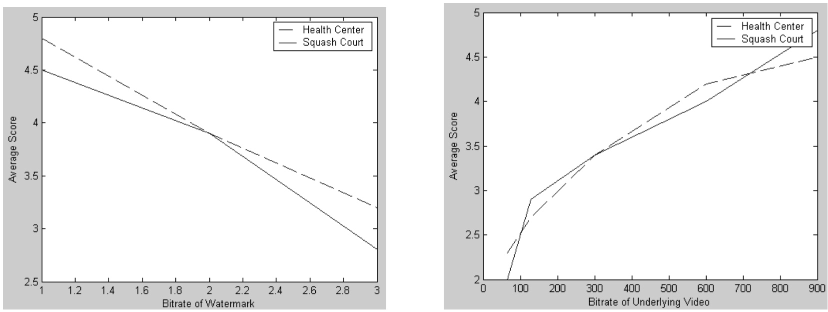
Fig. 4. a) Watermarked Video of bitrates 900 Kb/sec and Watermark payload rate of 3Kb/sec and without the Local adaptive gain module. b) Watermarked Video of bitrates 900 Kb/sec and Watermark payload rate of 3Kb/sec with Local adaptive gain applied. The difference is quality is quite visible in low texture regions like the sky and the wall

5 Subjective Evaluations

A protocol for evaluating the quality of watermarked video frames has been defined based on ITU-T recommendation P.910 [9] and ITU-R Rec. BT.500 [10]. These recommendations suggest different protocols based on the goal of the evaluations and availability of ground truth. In Absolute Category Rating (ACR) protocol, the images are shown and then the questions related to its quality are asked. The scale is discrete with values 1-5. In Degradation Category Rating (DCR)protocol, first the source is presented followed by system under test. The scale use is discrete impairment scale.

The ACR, DCR scales are not followed because they used discrete scales and hence take away the freedom to differentiate from the use. A protocol was designed in which the subject was shown the ground truth and system under test but the scale is continuous. In each test session the user was first made familiar with the interfaces. This was followed by a mock test session to check the familiarity. Then we had a test session which started with some stabilizing sequences. The scores were than normalized and mean scores were computed.

The subjective test was done on 5 subjects who were end-users and not experts in Image-processing. The scale on which the subjects voted was continuous. In the first session of the test the underlying video had a bit rate of 900 Kb/s. In second session the bit rates were changed and individual frames were shown. The watermark rate remained 1 Kb/frame. The following graphs (fig 5(a) and fig. 5(b)) show the performance of the two videos.



(a) Payload Bitrate Changed (b) Video Bitrates Changed

Fig. 5. a) The Subjective evaluations of the videos with payload bit-rate changed. b) The Subjective evaluations of the videos with video bitrates changed

6 Conclusion

A new technique for watermarking MPEG-4 bit-streams in compressed domain is proposed. The technique is not only robust to synchronization errors but also improves the quality of videos using local adaptive gain technique.

References

[1] Hartung, F., Girod, B.: Watermarking of uncompressed and compressed video. Signal Processing **66** (1998) 283–301
[2] Hartung, F.: Watermarking and fingerprinting of uncompressed and compressed video. Signal Processing **66** (1998) 283–301
[3] ISO: Information technology - generic coding of moving pictures and associated audio information. ISO/IEC 13818-2 (1994)
[4] Langelaar, G., R.Lagendijk: Optimal differential energy watermarking of dct encoded images and video. IEEE Transactions on image Processing **10** (2001) 148–158
[5] F. Jordan, M. K., Ebrahimi, T.: Proposal of a watermarking technique for hiding/retrieving data in compressed and decompressed video. ISO/IEC document JTC1/SC29/WG11 MPEG97/M2281 (1997)
[6] Hsu, C., Wu, J.: Hidden digital watermarks in images. IEEE Transactions on Image Processing **8** (1999) 58–68
[7] D. Nicholson, P. K., Delaigle, J.: Watermarking in mpeg4 context. European Conference on Multimedia Application Services and Techniques (1999) 472–492
[8] ISO: Information technology - coding of audio/video objects:video. ISO/IEC 14486-2 (October 1998)
[9] ITU-T: Methodology for subjective assesment of the quality of television pictures. Recommendation P.910 (1996)
[10] ITU-R: Subjective video quality assesment methods for multimedia applications. Recommendation BT. 500-11 (1995)